# LegoTherm: Modular and Reusable Thermal Model for Chiplet-based Heterogeneous Integration

Pengju Chen[1,2], Dan Niu[1,*], Depeng Xie[3], Gang Wang[4], Chen Wu[4], Wei W. Xing[5,*], Lei He[2,6,*]
[1]School of Automation, Southeast University, Nanjing, P.R. China
[2]Eastern Institute of Technology, Ningbo, P.R. China
[3]BTD Technology Inc., Nanjing, P.R. China
[4]Ningbo Institute of Digital Twin, Eastern Institute of Technology, Ningbo, P.R. China
[5]The University of Sheffield, Sheffield, United Kingdom
[6]University of California, Los Angeles, USA

## Abstract

Thermal management critically affects the reliability of chiplet-based heterogeneous integrated circuits (ICs) in the post-Moore era. The design process demands hundreds of thermal simulations across varying power distributions, cooling conditions, and structural configurations. Conventional methods based on finite element methods (FEM) and compact thermal model (CTM) lack the flexibility for efficient multi-scenario evaluations and incur prohibitive computational cost when modeling complex vertical stacks with cross-scale interconnects. We propose LegoTherm, a modular thermal modeling framework that decomposes chiplet-based systems into reusable reduced-order components. The framework exploits the inherent hierarchical structure of heterogeneous integration to construct high-fidelity 3D meshes capturing micro-interconnection details through modular discretization, then generates reduced-order modules by aggregating thermally coupled interface ports. This approach preserves critical hotspot accuracy while enabling rapid reassembly for different design scenarios. Evaluated on industrial-scale 2.5D and 3D packaging benchmarks, LegoTherm achieves up to 10.39× speedup compared to COMSOL while maintaining mean relative error below 0.40%. The framework reduces thermal design iteration cycles from weeks to half a day, addressing a critical bottleneck in modern IC design.

## Keywords

Chiplet, Thermal Modeling, Reusable Module, Reduced-order Model, Thermal Resistance Network

**ACM Reference Format:**

Pengju Chen[1,2], Dan Niu[1,*], Depeng Xie[3], Gang Wang[4], Chen Wu[4], Wei W. Xing[5,*], Lei He[2,6,*]. 2026. LegoTherm: Modular and Reusable Thermal Model for Chiplet-based Heterogeneous Integration. In *63rd ACM/1EEE Design Automation Conference (DAC'26), July 26-29, 2026, Long Beach, CA, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3770743.3803910

## 1 Introduction

As semiconductor manufacturing approaches fundamental physical limits, traditional Moore's law scaling faces increasingly severe challenges. The cost and complexity of monolithic system-on-chip (SoC) designs have risen sharply [1]. Chiplet-based heterogeneous integration has emerged as a pivotal pathway for enhancing system performance and energy efficiency beyond these scaling limits [2]. As illustrated in Figure 1, this architecture exhibits a hierarchical stacking structure analogous to Lego bricks. Multiple functionally diverse chiplets (dies) integrate through advanced packaging technologies and connect via micro-scale interconnections including through-silicon vias (TSVs), redistribution layers (RDLs), and $\mu$bumps (micro-bumps). This paradigm enables high-speed inter-die communication and IP-level functional module reuse, establishing a foundation for next-generation high-performance ICs [3, 4].

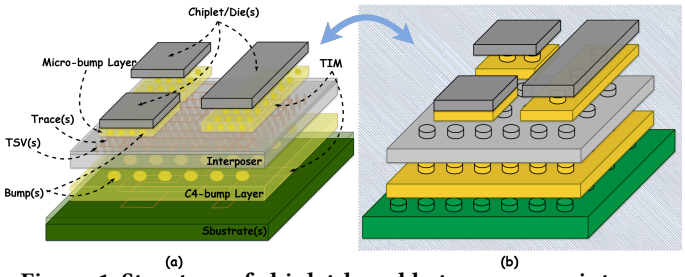



**Figure 1: Structure of chiplet-based heterogeneous integration. (a) 2.5D packaging with micro-scale interconnections. (b) Corresponding Lego structure mapping.**

The increasing complexity of hierarchical stacking significantly elevates power density in chiplet-based systems. Concentrated localized hotspots create steep thermal gradients that degrade device performance and reliability, making thermal management a critical step in the IC design process [5]. Traditional thermal simulation approaches fall into two categories [6]: analytical methods and numerical methods. Analytical methods, such as variable separation [7] and Green's function approaches [8, 9], remain confined to simple geometries with regular boundaries and cannot address complex heterogeneous structures.

Numerical methods discretize the domain through meshing and convert heat conduction partial differential equations into algebraic systems [10–13]. Among these, the finite element method (FEM) serves as the *Golden Standard* for accuracy and has been implemented in Manchester Thermal Analyzer [13] and commercial tools including COMSOL Multiphysics [14] and ANSYS Icepak [15]. However, FEM-based approaches require substantial computational resources when applied to chiplet-based heterogeneous integration with cross-scale features. Compact thermal models (CTM) based on finite difference methods (FDM), such as PTscalar [16], HotSpot [17], 3D-ICE [18], PACT [19], and the commercial tool Celsius [20], offer an alternative. These methods exploit electro-thermal duality to model chips as thermal resistance-capacitance (RC) networks, converting thermal simulation to circuit-level SPICE analysis. While CTM methods balance accuracy and efficiency, their reliance on

*Corresponding authors. This work is partially supported by NSFC (Grant No. 62374031), Natural Science Foundation of Jiangsu Province (BK20240173), NSFC (Grant No. 62331009, 62204265, 62234010, 92473107).

This work is licensed under a Creative Commons Attribution 4.0 International License.
*DAC'26, Long Beach, CA, USA*
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2254-7/2026/07
https://doi.org/10.1145/3770743.3803910

regularized mesh elements limits their ability to capture cross-scale interconnection features and localized hotspot details. Recently, deep learning approaches [21–23] have demonstrated significant speedup while maintaining acceptable accuracy. However, the feature matrix dimension increases sharply when modeling cross-scale interconnections, imposing high training costs and limiting scalability to complex chiplet-based designs.

A critical gap exists in current thermal modeling approaches. During actual system operation, chiplet-based heterogeneous integration must undergo repeated thermal simulation, verification, and optimization across multiple scenarios [24, 25]. These scenarios include variations in chiplet layout, stacking configurations, thermal interface materials (TIMs), power distributions, and cooling conditions. Existing methods focus exclusively on achieving accurate full-IC thermal simulation for a single configuration. Each modification requires reconstructing and re-solving the entire thermal model. This limitation is particularly acute for heterogeneous integration, *where chiplets are designed for reuse yet thermal tools require complete re-simulation for each combination.* Reduced-order models (ROM) offer a promising solution. ROM techniques have proven effective for analyzing complex systems in circuit and physics simulation [26–29], reducing system complexity while preserving accuracy. Applied to IC thermal modeling, ROM can map complex thermal structures to low-order equivalent models with reusable characteristics, reducing computational costs in multi-scenario analysis while preserving accuracy in critical regions.

We propose **LegoTherm**, a Lego-inspired modular thermal modeling framework that addresses this gap. LegoTherm decomposes complex IC structures into reusable modules and enables efficient multi-scenario simulations through module-level assembly. This framework significantly improves thermal analysis efficiency during the chiplet-based heterogeneous integration design phase. Our primary contributions include:

- **Lego Brick Construction**: We propose a hierarchical modular decomposition strategy enabling independent high-fidelity meshing and thermal-R extraction for each component in chiplet-based heterogeneous integration.
- **Lego Slot Aggregation**: We develop an aggregation scheme based on the thermal coupling relationships between module interface ports. The coupled ports are merged into supernodes to construct reusable reduced-order thermal modules while preserving inter-module energy conservation.
- **Lego Module Assembly**: We introduce modular assembly methods for fast reduced-order module reuse and simulation under different scenarios. It supports thermal evaluation and optimization across multiple design iterations.

## 2 Background and Problem Formulation

### 2.1 Thermal-R Network for Integrated Circuits

The principle of energy conservation in heat conduction closely resembles Kirchhoff's Current Law (KCL) in electrical circuits. Both enforce a balance of conserved quantities (heat flux $Q$ and current $I$) at each node $v$, ensuring total inflow equals total outflow [30]:

$$\sum_{i=1}^{n} I_{v_i} = 0 \quad \Leftrightarrow \quad \sum_{i=1}^{n} Q_{v_i} = 0. \tag{1}$$

This electro-thermal duality enables one-to-one mapping between electrical and thermal quantities, as summarized in Table 1. By exploiting this analogy, thermal analysis can leverage mature circuit simulation techniques.

**Table 1: Thermal–Electrical Analogy Mapping**

| Electrical Domain | Thermal Domain |
|---|---|
| Current Source | Heat Source |
| Voltage Source | Initial Condition and Ambient Condition |
| Current ($A$) | Rate of Heat Flow/ Power ($W$) |
| Voltage ($V$) | Temperature ($K$) |
| Resistance ($\Omega$) | Thermal Resistance ($K/W$) |
| Capacitance ($F$) | Thermal Capacitance ($J/K$) |

Thermal capacitance is relevant for transient thermal simulations and is not discussed in detail in this paper.

In thermal simulation of chiplet-based heterogeneous integration, the full chip can be discretized into multiple grids. For two adjacent grids $g_i^{(*)}$ and $g_j^{(*)}$ (where $*$ denotes the module index), the thermal resistance $R_{ij}$ between their centroids $v_i$ and $v_j$ is:

$$R_{ij} = \frac{l_i}{k_i \cdot A_{ij}} + \frac{l_j}{k_j \cdot A_{ij}}, \tag{2}$$

where $A_{ij}$ is the effective heat contact area, $l_i$ and $l_j$ are the distances from centroids to the contact interface, and $k_i$ and $k_j$ are the thermal conductivities of the two grids. For a grid element $g_i^{(*)}$ on the boundary with heat exchange area $S_i$, the ambient heat transfer resistance $R_{i_{\text{amb}}}$ is:

$$R_{i_{\text{amb}}} = \frac{l_{i_{\text{amb}}}}{k_i \cdot S_i} + \frac{1}{h \cdot S_i}, \tag{3}$$

where $l_{i_{\text{amb}}}$ is the distance from centroid $v_i$ to the heat exchange surface and $h$ is the convective heat transfer coefficient.

### 2.2 Problem Formulation

In steady-state thermal analysis, the global thermal equilibrium of chiplet-based heterogeneous integration satisfies:

$$\mathbf{G}\,\mathbf{T} = \mathbf{Q}, \tag{4}$$

where $\mathbf{G} \in \mathbb{R}^{N\times N}$ is the thermal conductance matrix, $\mathbf{T} \in \mathbb{R}^N$ is the temperature vector, and $\mathbf{Q} \in \mathbb{R}^N$ is the injected heat source vector. Consider the geometric domain of the heterogeneous IC as $\Omega \in \mathbb{R}^3$ and decompose it into $M$ independent modules satisfying $\Omega = \bigcup_{i=1}^{M} \Omega^{(i)}|_{\Omega^{(i)}\cap\Omega^{(j)}=\varnothing\ (i\neq j)}$. For each module $\Omega^{(i)}$, it interfaces with its neighboring modules through a set of interface ports $\mathbf{p}^{(i)}$, which serve as the sole thermal interaction channels.

Then, a reduced representation $\mathcal{R}^{(i)}$ is constructed for module $\Omega^{(i)}$. This representation preserves the input/output (I/O) thermal response at the $\mathbf{p}^{(i)}$ while suppressing the complexity of $\Omega^{(i)\prime}$s internal structure. Conceptually, $\mathcal{R}^{(i)}$ provides a compact operator:

$$\mathcal{R}^{(i)} : \mathbf{Q}_{\mathbf{p}^{(i)}} \mapsto \mathbf{T}_{\mathbf{p}^{(i)}}. \tag{5}$$

A number of $\mathbf{p}^{(i)}$ in $\mathcal{R}^{(i)}$ will also reduce its computational efficiency. Consequently, an additional port-aggregation step is necessary to lower the port dimensionality and enhance the overall computational performance:

$$\{p_1^{(i)}, p_2^{(i)}, ..., p_g^{(i)}\} \rightarrow \hat{\mathbf{p}}_g^{(i)} \in \mathbf{p}^{(i)}. \tag{6}$$

After that, the final reduced thermal module can be expressed as: $[\hat{\mathcal{R}}^{(i)}|\hat{\mathcal{R}}^{(i)} \leftarrow \hat{\mathbf{p}}_g^{(i)}]$. These reduced module models can be assembled according to the system's structural connectivity to approximate the global thermal solution:

$$\mathcal{T}\{\mathcal{F}_s\big[\hat{\mathcal{R}}^{(1)}, \hat{\mathcal{R}}^{(2)}, \ldots, \hat{\mathcal{R}}^{(M)}\big]\} = \hat{\mathbf{T}}_{\mathbf{s}}, \tag{7}$$

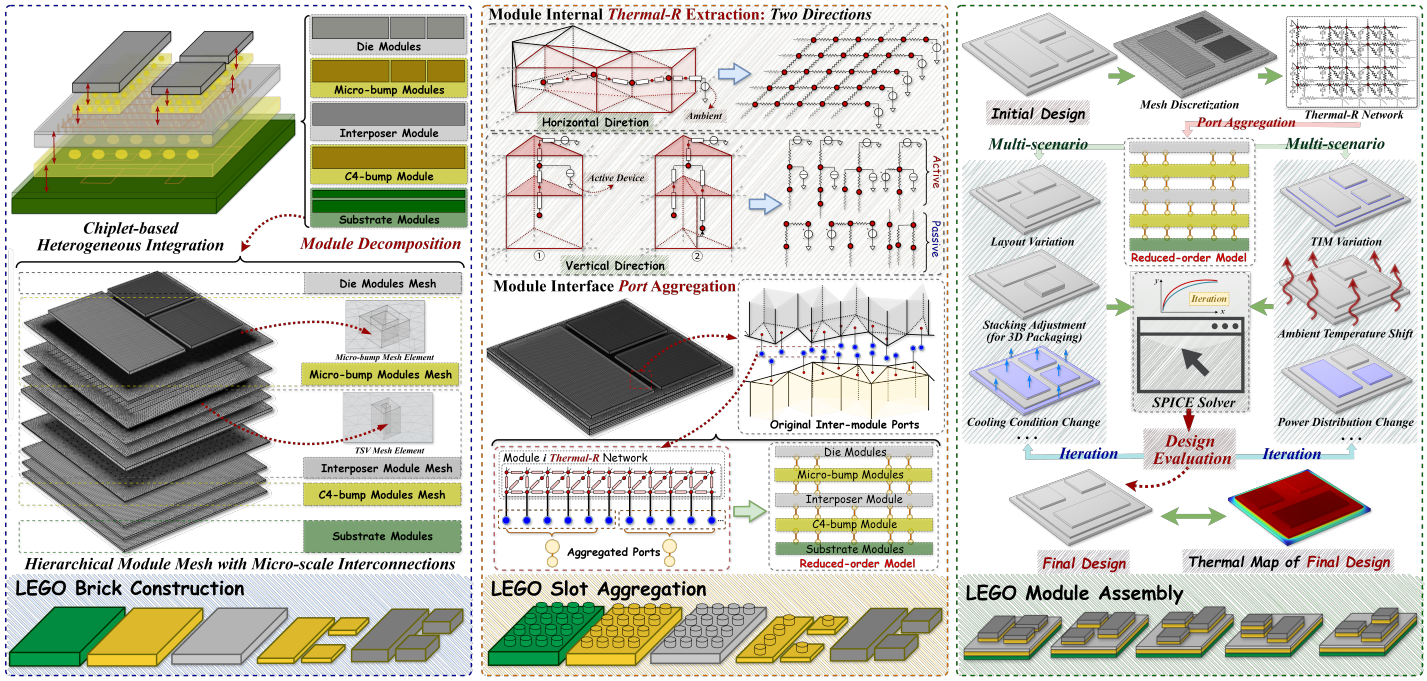



**Figure 2: The modular and reusable thermal modeling Framework LegoTherm for multi-scenario simulations.**

where $\mathcal{T}(\cdot)$ is thermal behavior and $\mathcal{F}_s(\cdot)$ is composition of modules under scenario $s$. The objective is to construct module-level reduced-order representations that support efficient reuse across multiple scenarios. Given a set of scenarios $\mathcal{S}$ representing different structural configurations, power distributions, and boundary conditions, the $\hat{\mathcal{R}}^{(i)}$ should enable rapid assembly and solution for each scenario $s \in \mathcal{S}$ while maintaining high accuracy:

$$\min_{\hat{\mathcal{R}}^{(i)}} \max_{s \in \mathcal{S}} \|\mathbf{T}_s - \hat{\mathbf{T}}_s\|_2 \leq \varepsilon, \tag{8}$$

where $\varepsilon$ is the acceptable error bound.

## 3 Proposed Methodologies

LegoTherm achieves efficient thermal simulation of chiplet-based heterogeneous integration under multiple scenarios by constructing reusable reduced-order modular thermal models. As illustrated in Figure 2, the framework comprises three key stages: (1) hierarchical decomposition of functional modules and mesh discretization, (2) thermal resistance extraction and reduced-order module generation through port aggregation, and (3) iterative thermal simulation based on reusable modules to obtain optimized IC thermal design.

### 3.1 Lego Brick Construction

Chiplet-based heterogeneous integration adopts a modular design approach, decomposing monolithic functionality and achieving spatial integration through multi-layer packaging and micro-scale interconnections. In the pre-processing for thermal simulation, the entire system is hierarchically decomposed into Lego brick-like modules based on function, such as die modules, $\mu$bump modules, interposer module, and so on. Each functional module $\Omega^{(i)} \subset \Omega$ can be further decomposed along the $Z$-axis into $L^{(i)}$ macroscopic geometric sub-layers according to material properties and functional characteristics:

$$\Omega^{(i)} = \bigcup_{l=1}^{L^{(i)}} \Omega_l^{(i)}, \quad \Omega_l^{(i)} \cap \Omega_k^{(i)} = \varnothing (l \neq k). \tag{9}$$

Each sub-layer isolates specific intra-module structures, such as TSV arrays, RDL layers, TIM layers, or bump layers, allowing tailored mesh density for cross-scale micro-interconnections. For each sub-layer $\Omega_l^{(i)}$, the 3D layer structures are orthogonally projected along the $Z$-axis to obtain its closed two-dimensional region $\Upsilon_l^{(i)}$ on the $XY$-plane:

$$\Upsilon_l^{(i)} = [P(\Omega_l^{(i)}) | z^{\text{top}}_{\Omega_l^{(i)}} \rightarrow z^{\text{bottom}}_{\Omega_l^{(i)}}], \tag{10}$$

where $P(\cdot)$ denotes the projection operator, $z^{\text{top}}_{\Omega_l^{(i)}}$ and $z^{\text{bottom}}_{\Omega_l^{(i)}}$ are the top and bottom elevations of sub-layer $l$. This projection captures the horizontal distribution of structures within the sub-layer. The Delaunay triangulation algorithm [31] is applied to generate a 2D triangular mesh over this region:

$$\mathcal{T}^{(i)}_{l(2D)} = \text{Delaunay}[\Upsilon_l^{(i)}]. \tag{11}$$

The Delaunay property ensures that the circumscribed circle of any triangle does not contain other vertices, guaranteeing good mesh quality and enabling fine mesh modeling of micro interconnections. After completing 2D meshing, the mesh is uniformly replicated and extruded along the $Z$-axis according to the sub-layer thickness $d_l^{(i)} = |z^{\text{top}}_{\Omega_l^{(i)}} - z^{\text{bottom}}_{\Omega_l^{(i)}}|$ to generate complete 3D volumetric elements. Let $z_l^{(i)}$ denote the number of divisions along the $Z$-axis. The $Z$-coordinate of each node is computed as:

$$\begin{aligned} z_n(\Omega_l^{(i)}) &= z^{\text{bottom}}_{\Omega_l^{(i)}} + (n - 0.5) \cdot \Delta z(\Omega_l^{(i)}), \quad n = 1, 2, ..., z_l^{(i)}, \\ \Delta z(\Omega_l^{(i)}) &= \frac{z^{\text{top}}_{\Omega_l^{(i)}} - z^{\text{bottom}}_{\Omega_l^{(i)}}}{z_l^{(i)}}. \end{aligned} \tag{12}$$

The complete 3D triangular prism mesh is then represented as:

$$\mathcal{T}^{(i)}_{(3D)} = \bigcup_{l=0}^{L^{(i)}} \text{Extrude}[\mathcal{T}^{(i)}_{l(2D)}, z_n(\Omega_l^{(i)})], \tag{13}$$

where Extrude[·] denotes the extrusion operation along the $Z$-axis. Through this approach, chiplet-based heterogeneous integration decomposes into independent Lego brick-like modules (without slots). Each module is discretized independently to meet its specific mesh granularity requirements while ensuring high-fidelity representation of micro-interconnection details.

## 3.2 Lego Slot Aggregation

Based on the electro-thermal duality, the thermal resistance network extracted from the hierarchical module mesh accurately characterizes heat conduction behavior within each module. Since interfaces between modules form critical heat transfer paths from heat sources to substrate in chiplet-based heterogeneous integration, the contact thermal ports between modules can be aggregated to form super-node port structures analogous to Lego slots.

*3.2.1 Module Internal Thermal Resistance* ***Extraction***. Heat transfer within a module can be quantitatively modeled via a thermal resistance network. The centroid of the discretized mesh element is treated as the node in this network. In horizontal direction, two adjacent elements $E_m$ and $E_n$ typically share one face $A_{mn}$. The thermal resistance between their centroids is calculated using Eq.(2). In vertical direction, an element $E_l$ may be thermally coupled with multiple elements $E_m$, $E_n$, etc., through multiple faces due to different functional sub-layers. And the thermal resistances between the centroid of $E_l$ and those of the adjacent elements based on the effective heat transfer areas of the corresponding faces also are calculated by Eq.(2) to establish the corresponding thermal connections.

For mesh elements located on the module surface, their centroids connect to an equivalent voltage source through boundary thermal resistance and convective thermal resistance as defined in Eq.(3) to simulate heat transfer boundary conditions. In active modules, equivalent current sources are introduced at the corresponding nodes according to the power density to represent heat power injection.

*3.2.2 Module Interface Port* ***Aggregation***. Vertically adjacent modules are thermally coupled through interfaces. Due to cross-scale micro-interconnections and the characteristics of Delaunay triangulation, each module includes numerous internal nodes and its interface typically contains a large number of original thermal coupling ports. This results in a high-order thermal conductance matrix during subsequent computations. To enable efficient reuse of thermal modules under multiple scenarios, we propose a novel model order reduction approach. By aggregating interface coupling ports to form super-node port structures analogous to Lego slots and reducing the number of internal nodes within each module. It significantly reduces computational complexity while preserves the I/O characteristics.

Let the set of heat transfer ports on the interface of module $\Omega^{(i)}$ be denoted as $\mathcal{P}^{(i)}$:

$$\mathcal{P}^{(i)} = \{P_1^{(i)}, P_2^{(i)}, ..., P_n^{(i)}\}, \tag{14}$$

where the port $P_k^{(i)}$ corresponds to a boundary node $v_k^{(i)}$, and there is a thermal resistance connecting $v_k^{(i)}$ with the centroid of the connected boundary element. These nodes $\mathbf{v}^{(i)}$ are distributed within the module according to material regions. When located in the same material region with geometric connectivity, their thermal behavior is physically similar, providing a reasonable basis for aggregating interface ports.

The thermal conductance matrix of module $\Omega^{(i)}$ can be expressed as:

$$\mathbf{G}^{(i)} = \begin{bmatrix} g_{11} & g_{12} & \cdots & g_{1I} & g_{1(I+1)} & \cdots & g_{1(I+B)} \\ g_{21} & g_{22} & \cdots & g_{2I} & g_{2(I+1)} & \cdots & g_{2(I+B)} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ g_{(I+B)1} & g_{(I+B)2} & \cdots & g_{(I+B)I} & g_{(I+B)(I+1)} & \cdots & g_{(I+B)(I+B)} \\ \underbrace{n_1}_{\text{Internal Nodes}} & n_2 & \cdots & n_I & \underbrace{n_{(I+1)}}_{\text{Boundary Nodes}} & \cdots & n_{(I+B)} \end{bmatrix}^{(i)}, \tag{15}$$

where the thermal conductance $g_{mn} = 1/R_{mn}$. Based on the geometric properties of the module's nodes, $\mathbf{G}^{(i)}$ can be expressed in block form as:

$$\mathbf{G}^{(i)} = \begin{bmatrix} \mathbf{G}_{ii} & \mathbf{G}_{ib} \\ \mathbf{G}_{bi} & \mathbf{G}_{bb} \end{bmatrix}^{(i)} \in \mathbb{R}^{(I+B)\times(I+B)}, \tag{16}$$

where $\mathbf{G}_{ii} \in \mathbb{R}^{I\times I}$ represents internal-internal coupling, $\mathbf{G}_{bb} \in \mathbb{R}^{B\times B}$ represents boundary-boundary coupling (zero for non-adjacent boundary elements), and $\mathbf{G}_{ib}$ and $\mathbf{G}_{bi}$ represent cross-coupling between internal and boundary nodes. By eliminating the internal degrees of freedom associated with steady-state heat distribution, the equivalent boundary thermal admittance is:

$$\mathbf{G}_{\text{am}}^{(i)} \in \mathbb{R}^{B\times B} = \mathbf{G}_{bb} - \mathbf{G}_{bi}\mathbf{G}_{ii}^{-1}\mathbf{G}_{ib}. \tag{17}$$

Some ports correspond to micro-interconnections that form critical heat-transfer paths (such as TSVs) and critical hotspots, they must be retained during aggregation. Accordingly, $\mathbf{G}_{\text{am}}^{(i)}$ is rearranged to separate aggregatable ports from those to be retained:

$$\hat{\mathbf{G}}_{\text{am}}^{(i)} = \begin{bmatrix} \mathbf{G}_{aa} & \mathbf{G}_{ap} \\ \mathbf{G}_{pa} & \mathbf{G}_{pp} \end{bmatrix}^{(i)}, \tag{18}$$

where $\mathbf{G}_{aa} \in \mathbb{R}^{A\times A}$ represents the relationships between aggregatable ports and $\mathbf{G}_{pp} \in \mathbb{R}^{(B-A)\times(B-A)}$ represents the relationships between microstructure ports that must be retained. The $A$ aggregatable ports can be grouped into $A'$ super-node sets based on their material properties and spatial continuity ($A \gg A'$). An aggregation mapping matrix $\mathbf{S}$ is defined to project the original port heat flux–temperature relationships onto the reduced-order super-node port space:

$$\mathbf{S} = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{B-A} \end{bmatrix} \in \mathbb{R}^{(B+A'-A)\times B}, \tag{19}$$

where $\mathbf{A} \in \{0,1\}^{A'\times A}$ is an indicator matrix representing super-node port sets relative to the original ports. If port $P_k^{(i)}$ belongs to set $r$, then $A_{r,k} = 1$; otherwise, $A_{r,k} = 0$. Applying $\mathbf{S}$ to $\hat{\mathbf{G}}_{\text{am}}^{(i)}$, the reduced module boundary admittance can be obtained:

$$\begin{aligned} \tilde{\mathbf{G}}^{(i)} &= \mathbf{S}\,\hat{\mathbf{G}}_{\text{am}}^{(i)}\,\mathbf{S}^\top \in \mathbb{R}^{(B+A'-A)\times(B+A'-A)} \\ &= \begin{bmatrix} \mathbf{A}\mathbf{G}_{aa}\mathbf{A}^\top & \mathbf{A}\mathbf{G}_{ap} \\ \mathbf{G}_{pa}\mathbf{A}^\top & \mathbf{G}_{pp} \end{bmatrix}^{(i)}. \end{aligned} \tag{20}$$

For the internal admittance of aggregated super-ports, if an aggregate set $\mathcal{G}_r$ contains a set of ports $\{P_j^{(i)}\}$, its corresponding admittance is $G_r^{\text{eq}} = \sum_{P_j^{(i)}\in\mathcal{G}_r} g_{P_j^{(i)}}$.

**Input-Output Equivalence:** For any module after order reduction, given an applied interface temperature vector $\hat{\mathbf{T}} \in \mathbb{R}^{(B+A'-A)}$, the resulting heat flow is:

$$\hat{\mathbf{Q}} = \tilde{\mathbf{G}}^{(i)}\,\hat{\mathbf{T}} = \mathbf{S}\,\hat{\mathbf{G}}_{\text{am}}^{(i)}\,\mathbf{S}^{\top}\hat{\mathbf{T}}. \tag{21}$$

Using $\mathbf{T} = \mathbf{S}^{\top}\hat{\mathbf{T}}$ to assign the temperature vector back to the original boundary nodes, it can obtain:

$$\hat{\mathbf{Q}} = \mathbf{S}\,\hat{\mathbf{G}}_{\text{am}}^{(i)}\mathbf{T} = \mathbf{S}\mathbf{Q}, \tag{22}$$

where $\mathbf{Q} = \hat{\mathbf{G}}_{\text{am}}^{(i)}\mathbf{T}$ is the heat flow at the original ports. This demonstrates that the I/O characteristics exhibited by the reduced-order module are equivalent to those of the original module. The proposed method aggregates interface ports to form Lego slot-like super-nodes, preserving thermal conduction accuracy while providing reusable low-order modules for subsequent multi-scenario thermal simulations.

## 3.3 Lego Module Assembly

After aggregating the super-nodes of interface ports, each module can be regard as a Lego module (with slots) that preserves the I/O characteristics and equivalent thermal behavior as the original structure. Different modules can be flexibly thermally coupled through their super-nodes, enabling rapid reconstruction of target structures under diverse scenarios and supporting efficient design thermal evaluation and iteratively optimization.

In thermal design of chiplet-based heterogeneous integration, different scenarios typically include chiplet layout and stacking configuration adjustment, modifications of ambient temperature and cooling boundary conditions, variations in TIM properties and changes in power distribution within active modules. For layout and stacking modifications, structural reconfiguration can be simply achieved by reconnecting the reduced-order modules through their interface super-nodes. Changes in ambient temperature or cooling boundaries affect only the equivalent voltage sources and thermal resistances. Variations in TIM performance only require updating the equivalent thermal conductance of the corresponding ports, while all other modules retain their reduced-order representations. Power-distribution updates in active modules are applied directly through adjustments to the equivalent current sources at the corresponding active nodes. Thus, scenario changes require updating only a small portion of modules, while the remaining reduced modules can be directly reused and reassembled. LegoTherm avoids solving the full-order integration system and enabling substantial efficiency improvement in multi-scenario thermal analysis.

Furthermore, for high-resolution thermal simulation requirements in critical areas, high-order models can be retained for the modules involved in these areas, while low-order modules are used for other areas. Since low-order modules retain the original I/O characteristics, accurate thermal distribution can be ensured for areas requiring high resolution.

# 4 Experimental Evaluation

We evaluate LegoTherm on industrial-scale chiplet-based heterogeneous integration benchmarks. We first describe the experimental setup, then assess efficiency by comparing simulation time against commercial tools under multi-scenario. Finally, we verify accuracy under different scenarios using COMSOL as the golden reference.

**Benchmarks:** We employ four representative industrial-grade chiplet-based heterogeneous integration benchmarks from Cadence reference designs, including two 2.5D packaging cases and two 3D packaging cases. Each vertical stack comprises functional substrates, C4-bumps layer, TSV and RDL-embedded interposer, $\mu$-bumps layer, and chiplets (dies). We iteratively modify these four benchmarks across multiple scenarios, and details are specified in Table 2.

**Methodology and Platform:** All experiments are conducted on a VMware Virtual Platform server with Intel(R) Xeon(R) Platinum 8490H CPU (1.90GHz, 196GB RAM, 32 CPU cores). We evaluate thermal simulation efficiency under multi-scenario settings by comparing total simulation time across COMSOL, a commercial Full-*RC* thermal simulator and LegoTherm. The Full-*RC* simulator and LegoTherm employ same SPICE solver to ensure fair comparison. The power map of chiplets are defined according to operating power density and applied to the corresponding thermal-R network nodes in form of equivalent current sources.

## 4.1 Multi-Scenario Efficiency

We comprehensively evaluate the efficiency of LegoTherm under multi-scenario by comparing total end-to-end thermal simulation time against COMSOL and Full-*RC* across 100 different scenarios of each benchmark design. These scenarios include chiplet layout and stacking variations, ambient temperature shifts, heat-transfer interface modifications, power redistribution and variations in TIM properties. The aggregated results are summarized in Table 2, where the first (pre-) simulation time is also listed separately.

LegoTherm requires an initial port aggregation to construct reusable reduced-order thermal modules, which results in its first simulation time longer than Full-*RC*. However, under multi-scenario requirements, it achieves significant acceleration by leveraging these modules. For relatively simple benchmark (Chiplet_2.5D_3), LegoTherm achieves 6.70× and 2.89× speedups over COMSOL and Full-*RC*. As structural complexity increases, its efficiency is further amplified. For the most complex benchmark (Chiplet_3D_12), LegoTherm reaching 10.39× faster than COMSOL and 8.41× than Full-*RC*. On average across all benchmarks, it delivers 7.93× acceleration over COMSOL and 5.33× compared to Full-*RC*. These results indicate that LegoTherm can shorten the thermal analysis cycle for chiplet-based heterogeneous integration from weeks to

**Table 2: Details of Benchmarks and Comparison of Different Thermal Simulation Methods**

| Heterogeneous Benchmarks | Details of Benchmarks | | | | | Time Consumption (*s*) | | | | | | Time Speedups | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Number | | | Minimum Scale ($\mu m$) | | First(Pre-) simulation | | | Multi-scenario simulations (*100 times*) | | | *vs* COMSOL | *vs* Full-*RC* |
| | Chiplets | Bumps | TSVs | Component | Pitch | COMSOL | Full-*RC* | **LegoTherm** | COMSOL | Full-*RC* | **LegoTherm** | | |
| Chiplet_2.5D_3 | 3 | 61 | 1,600 | 20 | 200.00 | 803.12 | 350.34 | 561.22 | 80,544.95 | 34,775.19 | 12,023.15 | 6.70× | 2.89× |
| Chiplet_2.5D_4 | 4 | 89 | 3,600 | 20 | 125.42 | 1,251.90 | 709.43 | 1,111.82 | 126,058.29 | 71,579.54 | 18,084.81 | 6.97× | 3.96× |
| Chiplet_3D_7 | 7 | 109 | 6,400 | 20 | 88.61 | 2,445.37 | 1,953.64 | 2,525.38 | 244,525.69 | 193,560.34 | 31,998.06 | 7.64× | 6.04× |
| Chiplet_3D_12 | 12 | 217 | 10,000 | 20 | 68.69 | 6,120.95 | 4,938.35 | 5,918.59 | 601,946.84 | 487,250.33 | 57,949.82 | **10.39×** | **8.41×** |
| Average | - | - | - | - | - | - | - | - | - | - | - | 7.93× | 5.33× |

Pengju Chen[1,2], Dan Niu[1,*], Depeng Xie[3], Gang Wang[4], Chen Wu[4], Wei W. Xing[5,*], Lei He[2,6,*]

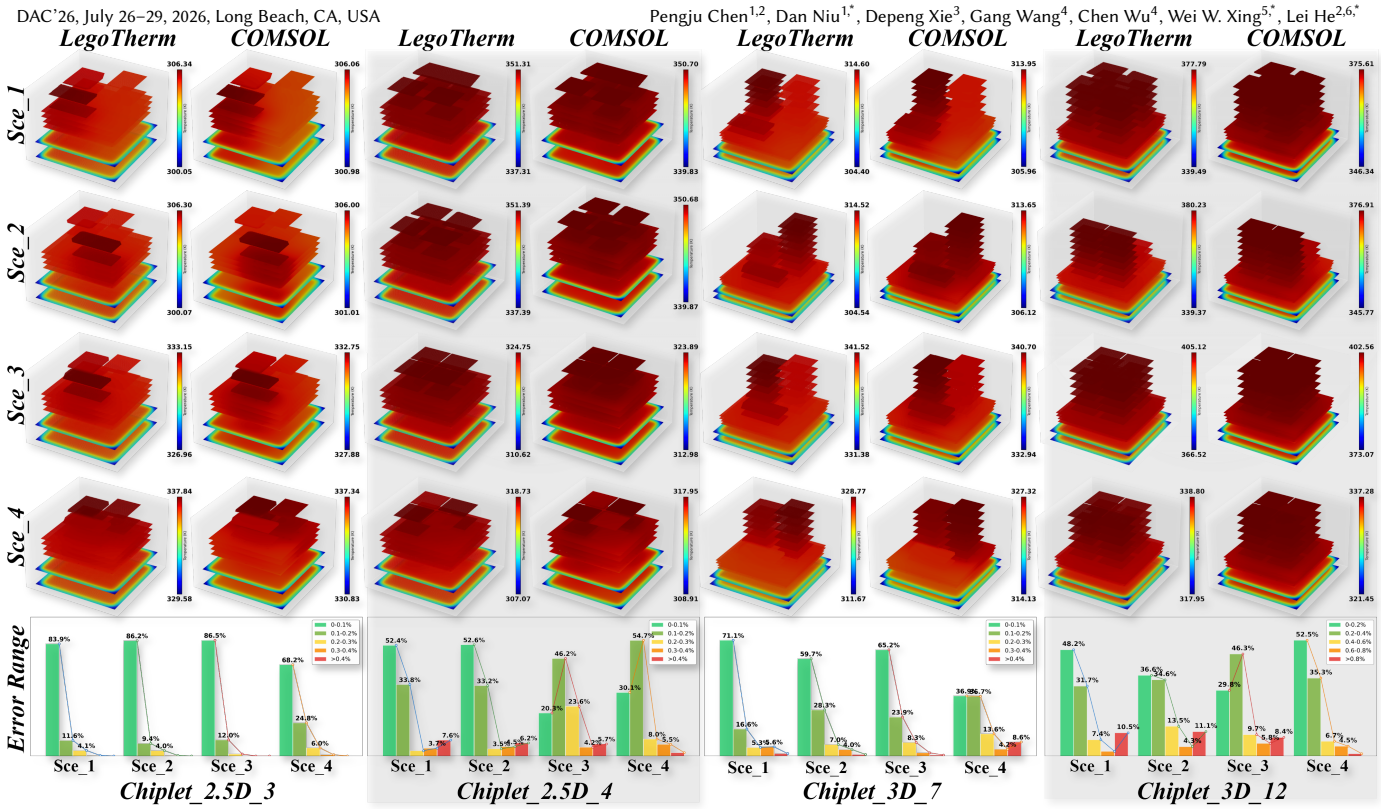



**Figure 3: Thermal maps and error ranges comparing LegoTherm against COMSOL across different benchmarks and scenarios.**

half a day, enabling far more efficient design iteration in modern IC packaging.

## 4.2 Multi-Scenario Accuracy

To evaluate the accuracy of the reduced-order thermal models reconstructed from reusable modules in LegoTherm, we use COMSOL as the golden reference. For each benchmark, we randomly select one reduced model assembled from reusable modules under four representative scenario types—layout variation (sce_1), stacking configuration changes (sce_2), ambient temperature shifts (sce_3), and power-distribution adjustments (sce_4). The corresponding error metrics for all scenarios are summarized in Table 3.

**Table 3: Thermal Errors of LegoTherm under Multi-Scenario**

| Heterogeneous Benchmarks | Multi-Scenario | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Sce.(s) | Error Metrics | | | Sce.(s) | Error Metrics | | |
| | | RMSE(K) | MAE(K) | MRE | | RMSE(K) | MAE(K) | MRE |
| Chiplet_2.5D_3 | Sce_1 | 0.26 | 0.20 | 0.06% | Sce_3 | 0.26 | 0.22 | 0.06% |
| | Sce_2 | 0.25 | 0.19 | 0.07% | Sce_4 | 0.36 | 0.30 | 0.09% |
| Chiplet_2.5D_4 | Sce_1 | 0.60 | 0.41 | 0.12% | Sce_3 | 0.66 | 0.57 | 0.17% |
| | Sce_2 | 0.61 | 0.47 | 0.13% | Sce_4 | 0.53 | 0.46 | 0.14% |
| Chiplet_3D_7 | Sce_1 | 0.41 | 0.28 | 0.09% | Sce_3 | 0.41 | 0.32 | 0.09% |
| | Sce_2 | 0.41 | 0.32 | 0.10% | Sce_4 | 0.68 | 0.52 | 0.16% |
| Chiplet_3D_12 | Sce_1 | 1.64 | 1.12 | 0.30% | Sce_3 | 1.69 | 1.35 | 0.34% |
| | Sce_2 | 1.71 | 1.32 | 0.36% | Sce_4 | 0.90 | 0.71 | 0.21% |

The quantitative results show that LegoTherm maintains an average relative error (MRE) below 0.2% in most scenarios. For more structurally complex benchmark (Chiplet_3D_12), the MRE increases slightly to 0.36% (scenario 4) but remains well within acceptable limits for thermal design. In addition, across all benchmarks and scenarios, the root-mean-square error (RMSE) ranges from 0.25 K to 1.71 K, while the mean absolute error (MAE) ranges from 0.19 K to 1.35 K.

Futhermore, to provide intuitive comparisons of spatial temperature profiles, we present the all-layer thermal maps produced by LegoTherm and COMSOL across all benchmarks under four scenarios and the corresponding error histograms in Figure 3. The results show that the reduced-order reusable model constructed with LegoTherm can effectively recover the temperature distribution of real chiplet-based heterogeneous integration. For most benchmarks (2.5D benchmarks and Chiplet_3D_7), more than 90% of spatial hotspots exhibit relative errors below 0.4%, with the majority falling below 0.2%. Even for structurally more complex benchmark (Chiplet_3D_12), it maintains high fidelity with more than 80% (84.7%) of locations staying within 0.6% relative error, and most points (at least 84.06%) remaining below 0.4%. These results confirm that LegoTherm preserves high accuracy while achieving significant acceleration.

## 5 Conclusion

This paper presents LegoTherm, a novel Lego-inspired modular and reusable thermal modeling framework that accelerates the thermal simulation of chiplet-based heterogeneous integration under multiple scenarios. By decomposing complex entire heterogeneous integration structures and aggregating interface ports to form reduced-order reusable modules, our approach achieves up to 10.39× speedup over COMSOL in total 100 different thermal simulation scenarios while maintaining MRE within 0.40% for complex 3D industrial-grade benchmark. Although LegoTherm requires more computation resources in the first simulation to construct reduced-order reusable modules, the limitation is outweighed by dramatic efficiency in subsequent multi-scenario iterations. LegoTherm enables rapid thermal–design iteration for modern ICs while maintaining accuracy, and can be effectively extended to dynamic thermal analysis because of the scalability of thermal-R network.

## References


[1] Yukai Chen, Venkateswarlu Sankatali, Subrat Mishra, Julien Ryckaert, James Myers, and Dwaipayan Biswas. 2024. Thermal implications in scaling high-performance server 3D chiplet-based 2.5 D SoC from FinFET to Nanosheet. In *2024 IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*. IEEE, 45–50.

[2] Xiaohan Ma, Ying Wang, Yujie Wang, Xuyi Cai, and Yinhe Han. 2022. Survey on chiplets: interface, interconnect and integration methodology. *CCF Transactions on High Performance Computing* 4, 1 (2022), 43–52.

[3] Chuanjun Nie, Qinzhi Xu, Chenghan Wang, He Cao, Jianyun Liu, and Zhiqiang Li. 2023. Efficient transient thermal analysis of chiplet heterogeneous integration. *Applied Thermal Engineering* 229 (2023), 120609.

[4] Yukai Chen, Massimiliano Di Todaro, Bjorn Vermeersch, Herman Oprins, Daniele Jahier Pagliari, Julien Ryckaert, Dwaipayan Biswas, and James Myers. 2025. Thermal Implications of Non-Uniform Power in BSPDN-Enabled 2.5 D/3D Chiplet-based Systems-in-Package using Nanosheet Technology. *arXiv preprint arXiv:2508.02284* (2025).

[5] Shunxiang Lan, Min Tang, and Junfa Mao. 2025. Hierarchical Decomposition and Interconnection Method for Efficient Thermal Simulation of Chiplet-Based 2.5 D System. In *2025 International Symposium of Electronics Design Automation (ISEDA)*. IEEE, 566–571.

[6] Liang Chen, Jincong Lu, Wentian Jin, and Sheldon X-D Tan. 2023. Fast full-chip parametric thermal analysis based on enhanced physics enforced neural networks. In *2023 IEEE/ACM International Conference on Computer Aided Design (ICCAD)*. IEEE, 1–8.

[7] Luqiao Yin, Ao Wang, Wenxing Zhu, Aiying Guo, Jingjing Liu, Min Tang, Liang Chen, and Jianhua Zhang. 2024. A Stepwise Integration Separation of Variables Solver for Full-Chip Thermal Uncertainty Analysis. *IEEE Transactions on Components, Packaging and Manufacturing Technology* 14, 4 (2024), 630–640.

[8] Yong Zhan and Sachin S Sapatnekar. 2007. High-efficiency Green function-based thermal simulation algorithms. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 26, 9 (2007), 1661–1675.

[9] Hameedah Sultan and Smruti R Sarangi. 2020. A fast leakage-aware Green's-function-based thermal simulator for 3-D chips. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* 28, 11 (2020), 2342–2355.

[10] Yousef Safari, Adam Corbier, Dima Al Saleh, Fahad Rahman Amik, and Boris Vaisband. 2025. Thermal Simulator for Advanced Packaging and Chiplet-Based Systems. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* (2025).

[11] Weijun Zhu, Gang Dong, and Yintang Yang. 2018. Thermal-aware modeling and analysis for a power distribution network including through-silicon-vias in 3-D ICs. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 38, 7 (2018), 1278–1290.

[12] Qipan Wang, Tianxiang Zhu, Yibo Lin, Runsheng Wang, and Ru Huang. 2024. ATSim3D: Towards Accurate Thermal Simulator for Heterogeneous 3D-IC Systems Considering Nonlinear Leakage and Conductivity. In *2024 2nd International Symposium of Electronics Design Automation (ISEDA)*. IEEE, 618–623.

[13] Scott Ladenheim, Yi-Chung Chen, Milan Mihajlović, and Vasilis F Pavlidis. 2018. The MTA: An advanced and versatile thermal simulator for integrated systems. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 37, 12 (2018), 3123–3136.

[14] Roger W Pryor. 2009. *Multiphysics modeling using COMSOL®: a first principles approach*. Jones & Bartlett Publishers.

[15] Stephen Scampoli. 2009. ANSYS Icepak®. *ADVANTAGE* (2009).

[16] Weiping Liao, Lei He, and K.M. Lepak. 2005. Temperature and supply Voltage aware performance and power modeling at microarchitecture level. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 24, 7 (2005), 1042–1053. doi:10.1109/TCAD.2005.850860

[17] Wei Huang, Shougata Ghosh, Sivakumar Velusamy, Karthik Sankaranarayanan, Kevin Skadron, and Mircea R Stan. 2006. HotSpot: A compact thermal modeling methodology for early-stage VLSI design. *IEEE Transactions on very large scale integration (VLSI) systems* 14, 5 (2006), 501–513.

[18] Arvind Sridhar, Alessandro Vincenzi, David Atienza, and Thomas Brunschwiler. 2013. 3D-ICE: A compact thermal model for early-stage design of liquid-cooled ICs. *IEEE Trans. Comput.* 63, 10 (2013), 2576–2589.

[19] Zihao Yuan, Prachi Shukla, Sofiane Chetoui, Sean Nemtzow, Sherief Reda, and Ayse K Coskun. 2021. PACT: An extensible parallel thermal simulator for emerging integration and cooling technologies. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 41, 4 (2021), 1048–1061.

[20] Neil Tyler. 2019. Cadence Launches Celsius Thermal Solver.

[21] Maziar Raissi, Paris Perdikaris, and George E Karniadakis. 2019. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics* 378 (2019), 686–707.

[22] Ziyue Liu, Yixing Li, Jing Hu, Xinling Yu, Shinyu Shiau, Xin Ai, Zhiyu Zeng, and Zheng Zhang. 2023. DeepOHeat: operator learning-based ultra-fast thermal simulation in 3D-IC design. In *2023 60th ACM/IEEE Design Automation Conference (DAC)*. IEEE, 1–6.

[23] Liang Chen, Wenxing Zhu, Min Tang, Sheldon X-D Tan, Jun-Fa Mao, and Jianhua Zhang. 2024. PISOV: Physics-Informed Separation of Variables Solvers for Full-Chip Thermal Analysis. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* (2024).

[24] Lukas Pfromm, Alish Kanani, Harsh Sharma, Parth Solanki, Eric Tervo, Jaehyun Park, Janardhan Doppa, Partha Pratim Pande, and Umit Ogras. 2024. MFIT: Multi-fidelity thermal modeling for 2.5 D and 3D multi-chiplet architectures. *ACM Transactions on Design Automation of Electronic Systems* (2024).

[25] Emad Haque, Pragnya Nalla, Chetan Choppali Sudarshan, Divya Yogi, Hangyu Zhang, Chaitali Chakrabarti, Vidya A. Chhabria, Ramesh Harjani, Jeff Zhang, and Sachin S. Sapatnekar. 2025. Invited: EDA for Heterogeneous Integration. In *2025 62nd ACM/IEEE Design Automation Conference (DAC)*. 1–4. doi:10.1109/DAC63849.2025.11132516

[26] Zizheng Guo, Tsung-Wei Huang, Zhou Jin, Cheng Zhuo, Yibo Lin, Runsheng Wang, and Ru Huang. 2024. Heterogeneous static timing analysis with advanced delay calculator. In *2024 Design, Automation & Test in Europe Conference & Exhibition (DATE)*. IEEE, 1–6.

[27] Shunxiang Lan, Min Tang, and Jun Ma. 2025. Flexible Thermal Conductance Model (TCM) for Efficient Thermal Simulation of 3-D ICs and Packages. In *2025 Design, Automation & Test in Europe Conference (DATE)*. IEEE, 1–6.

[28] George Floros, Nestor Evmorfopoulos, and George Stamoulis. 2019. Efficient IC hotspot thermal analysis via low-rank Model Order Reduction. *Integration* 66 (2019), 1–8.

[29] Tianxiang Zhu, Qipan Wang, Yibo Lin, Runsheng Wang, and Ru Huang. 2025. MORE-Stress: Model Order Reduction based Efficient Numerical Algorithm for Thermal Stress Simulation of TSV Arrays in 2.5 D/3D IC. In *2025 Design, Automation & Test in Europe Conference (DATE)*. IEEE, 1–7.

[30] Swapnil S Salvi and Ankur Jain. 2021. A review of recent research on heat transfer in three-dimensional integrated circuits (3-D ICs). *IEEE Transactions on Components, Packaging and Manufacturing Technology* 11, 5 (2021), 802–821.

[31] Yahia S Elshakhs, Kyriakos M Deliparaschos, Themistoklis Charalambous, Gabriele Oliva, and Argyrios Zolotas. 2024. A comprehensive survey on Delaunay triangulation: applications, algorithms, and implementations over CPUs, GPUs, and FPGAs. *IEEE Access* 12 (2024), 12562–12585.